DOI: 10.1007/s11676-007-0028-6

The RACE amplifying and sequence analyzing of secreted aspartic proteinase gene SA76 of *Trichoderma harzianum*

LIU Yan and YANG Qian *

Department of Life Science and Engineering, Harbin Institute of Technology, Harbin 150001, P.R.China

Abstract: Total RNA was isolated from mycelium of *T. harzianum* by Total RNA extraction kit, and two clear bands of rRNA (28S and 18S) were observed in agarose electrophoresis. By joining the 3′end sequence with the known SA76 EST from cDNA library of *T. harzianum*, a full-length cDNA sequence of 2019bp was obtained, whose open reading frame contained 1593bp, a stop codon TAA, a 5′untranslated region (5′UTR) of 266bp, a 3′untranslated region (3′UTR) of 201bp, and poly (A) 29 encoded a protein of 530 amino acids, had a signal peptide. *T. harzianum* shared 53% identity of secreted aspartic proteinase gene with *G. zeae*, 37% with *N. crassa* and 36% with *C. globosum*. The full-length cDNA sequence of secreted aspartic proteinase gene from *T. harzianum* was cloned for the first time by using BD SMART RACE technique, which provides a foundation to obtain and validate functional genes of *T. harzianum*.

Keywords: BD SMART RACE; *T. harzianum*; secreted aspartic proteinase gene

Introduction

T. harzianum species are commercially applied as biological control agents against a large number of plant fungal pathogens with different mechanisms, such as the production of antifungal metabolites, competition for space and nutrients and mycoparasitism (Harman and Björkman 1998; Manczinger and Polner 1985). Another activity associated with biocontrol function can be used with a protease to inactivate some enzymes that are secreted by pathogenic fungi. It prevents effectively fungal pathogens from penetrating the plant organisms. For instance, a hydrolytic enzyme playing an important role in mycoparasitism was purified and biochemically characterized by serine-protease (Geremia *et al*. 1993). However, very few studies were related to the protease expression by antipathogenic fungi (Nathalie *et al*. 2001).

To explore the mechanism of biocontrol agents, a cDNA library from *T. harzianum* mycelium was constructed and 3298 ESTs were acquired after sequencing (Liu and Yang 2005). A partial cDNA sequence of secreted aspartic proteinase gene (SA76) has been obtained by step-walking sequencing. Rapid amplification of cDNA ends (RACE) is one of the main methods for extending partially known exon sequence and cloning the full-length cDNA (Schaefer 1995). It has been widely used in further extension of functional fragments in genes of interest especially the cloning of the full-length cDNAs. Acquisition of full-length cDNA from EST is important to genomics and functional genomics (Louie and Patricia 1999; Zheng and James 2000). RACE is one of research ways in this field. BD SMART RACE is usually used to amplify the 3' ends of SA76.

In this study, the full-length SA76 gene was cloned firstly from mycelium of *T. harzianum* by BD SMART RACE technique, which provided the foundation for further study on structure and function of SA76. Moreover, we can further know the model of secreted aspartic proteinase hydrolytic pathway and its bio-control mechanism.

Materials and methods

Materials

All reagents were of the highest purity available commercially. AdvantageR® 2 PCR Enzyme System®, SMART™ PCR cDNA Synthesis Kit, BD SMART™ RACE cDNA Amplification Kit and Nucleo Trap® Nucleic Acid Purification Kits were from Clontech (Clontech Laboratories, Palo Alto,CA). TA cloning Kit from TaKaRa. Total RNA extracted kit from Watson (Watson Biotechnologies, INC)

Methods

Total RNA extraction

Total RNA was extracted from mycelium of *T.harzianum* using Total RNA extracted kit (Watson).

Primers of 3' RACE

3'-RACE CDS Primer A :

5'–AAGCAGTGGTATCAACGCAGAGTAC$(T)_{30}$V N–3'

(N = A, C, G, or T; V = A, G, or C)

10X Universal Primer A Mix (UPM):

Foundation project: This paper was supported by Development Program (863) of China (No. 2003AA241140)

Received date: 2006-12-23; Accepted date: 2007-02-26

Electronic supplymentary material is available in the online version of this article at http://dxdoi.org/10.1007/s11676-007-0028-6

Biography: LIU Yan (1976-), female, Ph.D Candidate, in the department of Life Science, Harbin Industrial University, Harbin 150001, P. R. China. Email: liuyanyan626@163.com

*Corresponding author: YANG Qian (E-mail: liuyanyan626@163.com)

Responsible editor: Chai Ruihai

Long:
5'–CTAATACGACTCACTATAGGGCAAGCAGTGGTATCAA CGCAGAGT–3'
Short: 5'–CTAATACGACTCACTATAGGGC–3'
Nested Universal Primer A (NUP):
5'–AAGCAGTGGTATCAACGCAGAGT–3'
GSP3: 5'CCAACCGTACAGCTTGCTGCTCAACAC3'
NGSP3: 5'AGCCCTCATCTATGGCGGTCTTGATCGCAG3'

Primers of verified RACE results
F1: 5'CCCAAGCTTTCTTCCGTCTCTTTGTTCCTCTTTC3'
R1: 5'CCGGAATTCATGGTCCTCTCGCTCGTCAATATAT3'

RACE PCR and cloning the full-length cDNAs
The 3' ends of the transcripts were amplified by rapid amplification of the cDNA ends (RACE) using the BD SMART^{TM} RACE cDNA Amplification Kit according to the manufacturer's instructions (Clontech). The double-stranded cDNAs were prepared from total RNAs of mycelium from *Trichoderma harzianum*. Primers were designed according to the sequences of cDNA clones above. The first PCR reaction was carried out in a Gene Amp® PCR system 9700, 'Touchdown'

PCR cycling conditions were 94ºC, 30s, 72 ºC, 3 min, 6 cycles; 94 ºC, 30 s, 70 ºC, 30s, 72 ºC, 3 min, 6 cycles; 94 ºC, 30 s, 68 ºC, 30s, 72 ºC, 3 min 25 cycles. In the second PCR reaction, an aliquot of the primary PCR product is reamplified using the inner primers. Conditions for PCR amplification were 94 °C for 2 min; 30 cycles at 94°C for 30 s, 68°C for 30 s and 72°C for 2 min; followed by a final extension at 72 ºC for 10 min.

PCR products were analyzed by agarose gel electrophoresis, cloned using the TA cloning kit. The inserts were sequenced.

Results

Total RNA isolation

Total RNA was isolated from mycelium of *T. harzianum* by Total RNA extraction kit, and two clear bands of rRNA (28S and 18S) were observed in agarose electrophoresis (Fig. 1). The brightness of 28S rRNA was nearly 2 times that of 18S rRNA. In addition, the ratio of OD260/OD280 was 1.98. Result showed that the RNA was very little degraded and the purity of RNA was high. Moreover, it also suggested that total RNA had been used successfully in reverse transcription of *T. harzianum* cDNAs.

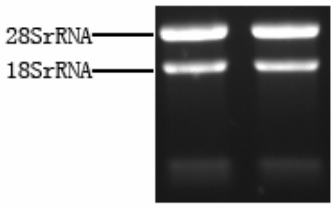

Fig. 1: Total RNA from *T. harzianum*

Cloning of SA76 gene by BD SMART RACE Technique

The results of 3'RACE
The 1 % agarose gel electrophoresis for the first PCR products showed no specific band but a smear was found (Fig. 2, Lane1). Two bands of about 1200bp and 700bp were showed in the electrophoresis for the second PCR products using inner primer (Fig. 3, Lane 1). The band of 1200bp was target sequence.

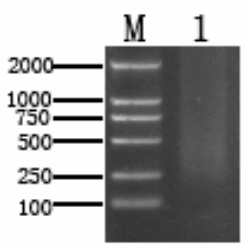

Fig. 2: 3′RACE product of the first PCR
Lane M: DL2000 DNA marker (TaKaRa ,Japan); Lane 1: 3′RACE product of the first PCR amplifications

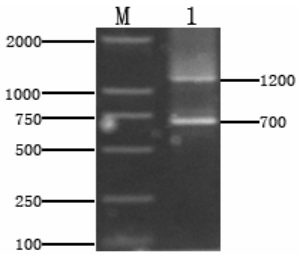

Fig. 3: 3′RACE product of the second PCR
Lane M: DL2000 DNA marker; Lane 1: 3′RACE product of the of the second PCR amplifications

Cloning of full-length cDNA
Sequencing results showed that 3′RACE products were 1233bp excluding vector sequence. A full-length cDNA of 2019bp was produced by joining 3′fragments with the known SA76 EST. The specific primers of 5′ and 3′ gene were designed to verify conjugation product. Analysis of 1 % agarose gel electrophoresis showed that the resulting fragment was approximately 1745bp in length (Fig. 4). It suggested that the full-length cDNA had been obtained successfully.

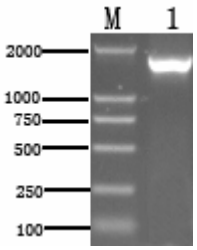

Fig. 4: verified the full-length cDNA with specific primers
Lane M: DL2000 DNA marker (TaKaRa, Japan); Lane 1: PCR product.

Sequence analysis of full-length cDNA
By joining the 3′end sequence from cDNA library of *T. harzianum*, a full-length cDNA sequence of 2019bp was acquired (Fig. 5), whose open reading frame contained 1593bp, a stop codon TAA, a 5′untranslated region (5′UTR) of 226bp, a 3′untranslated region (3′UTR) of 201bp and poly (A) 29 encoding a protein of 530 amino acids, with a theoretical molecular weight of 55.4kD and a calculated pI of 4.35. The nucleotide sequence data has been submitted to GenBank (accession number: EF063645).

```
1       TCGAATTCCGACGTTCCCCCGCTAGTGAGATCGCGCCGAGATCTGGGTTAATCACTCTGT
61      TGACTCGATTGCCTTCAGCTGGGACTCGATTTTCCTCTTTCTCTTTTCTCTAGCTCCCCC
121     CTTCTTCCGTCTCTTTGTTCCTCTTTCCACCCGCCGGAACCGATATCGTTGCGAACACAG
181     GGAGACAGAGAGACGAGAAGGAGAGTCGACAGAACCTTCGCACACCATGAGGCTGACGAC
                                                    M  R  L  T  T
241     AGCGCTCGGGCTGCTGATCGCGGCGCAACATGCGGAAGCCGTTGTCACTCCGATGTTCCC
          A  L  G  L  L  I  A  A  Q  H  A  E  A  V  V  T  P  M  F  P
301     GCGGGCGGAATCTGGTGATGGATATTTGTCGATCCCCGTGGGAACCATCAAGAGGCCTCA
          R  A  E  S  G  D  G  Y  L  S  I  P  V  G  T  I  K  R  P  H
361     CAACAAGGTTGGAAAGAGAAGCGCCATTGACGCAGTATTGGAGAATATGGATTTCTTCTA
          N  K  V  G  K  R  S  A  I  D  A  V  L  E  N  M  D  F  F  Y
421     TGCCATCGAAATCGGCCTAGGAACTCCTCCCCAGAACGTAACTGTCCTCGTCGATACAGG
          A  I  E  I  G  L  G  T  P  P  Q  N  V  T  V  L  V  D  T  G
481     ATCCAGCGAGCTATGGGTCAATCCGGACTGCTCGACCGCACCGTCCGAGTCGCAGGCCGA
          S  S  E  L  W  V  N  P  D  C  S  T  A  P  S  E  S  Q  A  E
541     ACAGTGTCAGCAGCTCGGCCAATACAATCCCAGAAGATCGAGAACGCCGCCGGTTGGTCC
          Q  C  Q  Q  L  G  Q  Y  N  P  R  R  S  R  T  P  P  V  G  P
601     GTTTGGACGCGAGGAAATCAACTATGGCGACCCAACAGACCAGTCCACGCAGACGTCAGT
          F  G  R  E  E  I  N  Y  G  D  P  T  D  Q  S  T  Q  T  S  V
661     CGACATCACCTACTATGCCGACACGCTGAGCTTTGGTAGGAGTCAGGTCAAGAATCAGAC
          D  I  T  Y  Y  A  D  T  L  S  F  G  R  S  Q  V  K  N  Q  T
721     GTTTGGCGTTGTTACGTCCAGCGAGGGCCAGGCACAGGGCATCATGGGCCTCGCGCCTGA
          F  G  V  V  T  S  S  E  G  Q  A  Q  G  I  M  G  L  A  P  D
781     TGTTCGAGGAGGATTTCCAGGCGACCAACCGTACAGCTTGCTGCTCAACACAATGGCCGA
          V  R  G  G  F  P  G  D  Q  P  Y  S  L  L  L  N  T  M  A  D
841     CCAGGGAGTCATTGCCAGCCGGGTCTTTTCCCTTGACCTCCGGCATTCCGATTCAGAGAC
          Q  G  V  I  A  S  R  V  F  S  L  D  L  R  H  S  D  S  E  T
901     GGGAGCCCTCATCTATGGCGGTCTTGATCGCAGCAAATTCATCGGTTCCCTCGAGACTCG
          G  A  L  I  Y  G  G  L  D  R  S  K  F  I  G  S  L  E  T  R
961     ACCCATCGTACCCGGCATCCAAGGCGAAACACGTCTGGCCGTAAATCTGACTACACTGGG
          P  I  V  P  G  I  Q  G  E  T  R  L  A  V  N  L  T  T  L  G
1021    CCTCACGCAAAGCCGTTCGCAGAGCTTCAGGCTGAACAGCGCCGACACAAACGTGATGCT
          L  T  Q  S  R  S  Q  S  F  R  L  N  S  A  D  T  N  V  M  L
1081    CGACTCTGGCACGACGCTCAGCCGCATGCACTCCGCTGCCGCATCGCCTATCCTCGAGAC
          D  S  G  T  T  L  S  R  M  H  S  A  A  S  P  I  L  E  T
1141    TCTGGGCGCCCAAAACGATGGCGAGGGCTACTTTTTTGTGCCGTGCTCGCTTCGTGACTC
          L  G  A  Q  N  D  G  E  G  Y  F  F  V  P  C  S  L  R  D  S
1201    CGCTGGCAGTGTCGATTTTGGCTTCGGCAACAAGGTCATCAGGGTTCCCTTCTCTGATTT
          A  G  S  V  D  F  G  F  G  N  K  V  I  R  V  P  F  S  D  F
1261    CATCCTATCAGCAGGCGATAGTGGCGGCCCCAGCGACTATTGTTATGTTGGCCTGGTCCT
          I  L  S  A  G  D  S  G  G  P  S  D  Y  C  Y  V  G  L  V  L
1321    GACGACGGACCAGCAGATTCTGGGAGACACGGTGTTGAGGGCTGGATACTTTGTATTCGA
          T  T  D  Q  Q  I  L  G  D  T  V  L  R  A  G  Y  F  V  F  D
1381    TTGGGACAACCAGGAGGTTCACATCGCCCAGGCCGCTGACTGTGGCAGCAGTGACATTGT
          W  D  N  Q  E  V  H  I  A  Q  A  A  D  C  G  S  S  D  I  V
1441    TGTCGCCGGCAGCGGATCCAAGGCCGTGCCCAATGTGCAGGGCAATTGCAACAGCAGTGA
          V  A  G  S  G  S  K  A  V  P  N  V  Q  G  N  C  N  S  S  D
1501    TGCCGGTGTCACGGGCACAGGAGGCCCAACGGCCACGGGATCGACCCCGACTCCCACCAA
          A  G  V  T  G  T  G  G  P  T  A  T  G  S  T  P  T  P  T  N
1561    CAATATCCCAGCTACAGCCGTTACAACTGTCTTCACGGTGACATCCTGCCCAGTTTTCGA
          N  I  P  A  T  A  V  T  T  V  F  T  V  T  S  C  P  V  F  D
1621    TGTCGGATGCCGCACCGGCATGATCACAACCCAAACCATTCAAGGAGCTGAAGCTACGCC
          V  G  C  R  T  G  M  I  T  T  Q  T  I  Q  G  A  E  A  T  P
1681    TCAACCCAGCGCTACCAGCACTCCTAGTAGCGGCGGCAATGGAGATGGAGGAGACGAAGA
          Q  P  S  A  T  S  T  P  S  S  G  G  N  G  D  G  G  D  E  D
1741    TGCAGGCGTGCGGCCTCCGGCCTTGACTTGGGTCTTTGTCGCGCTGGGTACTTTGGCAAT
          A  G  V  R  P  P  A  L  T  W  V  F  V  A  L  G  T  L  A  M
1801    GATTTTTAACATTGTATAATTTCAAGAAGGGGTCTCTTACTGTATATATTGACGAGCGAG
          I  F  N  I  V  *
1861    AGGACCATTGGCCATCAATTTTGGATAACTCGGCCGAGGATTTACATACCAGGTTCGTAT
1921    ATGACGATGGAGGGTGGGACATAGCTGCGCATGGAGGATTTATTTTCCCAATGGATACCT
1981    GACATACATAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
```

Fig. 5 Nucleotide sequence of cDNA and its deduced amino acid sequence.

Analysis of the amino acid sequence by SignalP v3.0 identified the existence of signal-sequence site, it was inferred from this result that it was a kind of secreted protein. Two possible cleavage sites of the signal sequence could be detected between amino acid position18 and 19(A, V) (Fig. 6).

BlastP analysis against non-redundant protein sequence database revealed that there was a high degree of similarity between SA76 and secreted aspartic proteinase belonging to the Eukaryotic aspartyl protease family (Fig.7). Homologous analysis of the deduced amino acid sequences was performed by BLASX and subsequently compared with GenBank data. Result indicated that the 530 amino acid residues of *G. zeae* shared 53% identity of secreted aspartic proteinase gene, and, the shared identities of *N. crassa* and *C. globosum* were 37% and 36%, respectively. Alignment of SA76 sequence with the closest protease allowed us to identify the residues that were necessary for a functionally active gene.

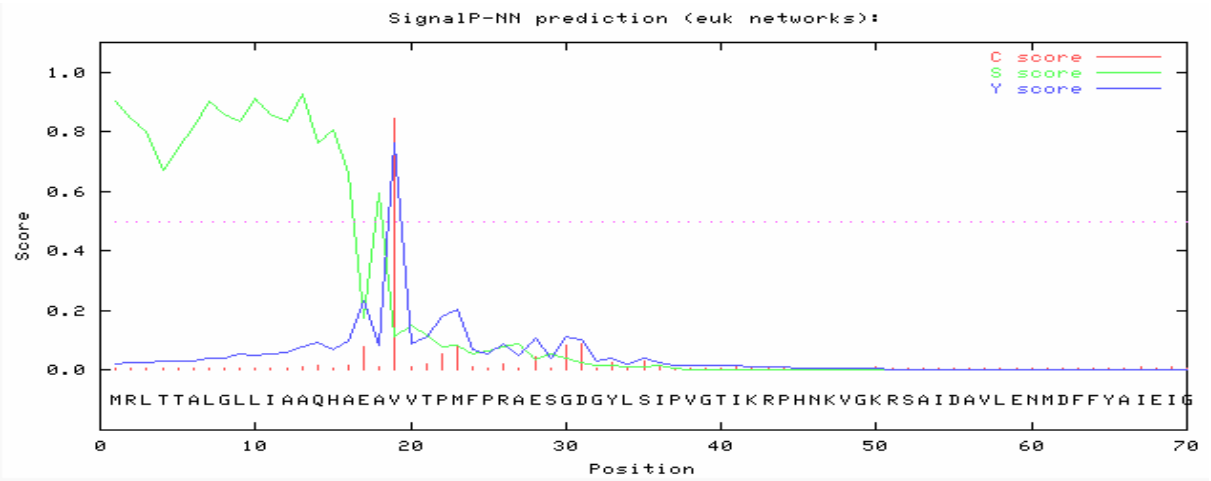

Fig.6 Signal peptide prediction

Fig. 7 Conserved domain prediction of SA76 gene

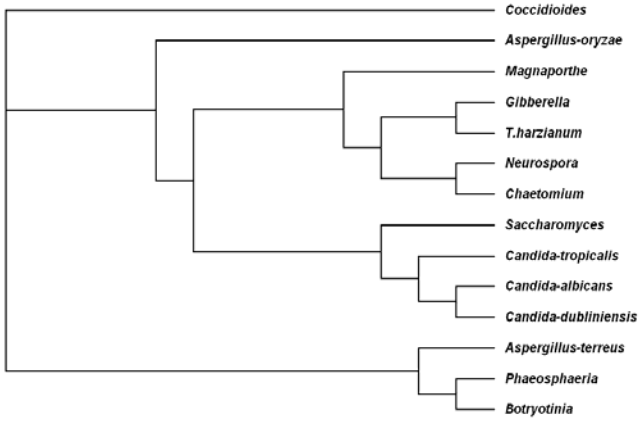

Fig. 8 Phylogenetic tree of SA76 gene from *T. harzianum*

As shown in Fig. 8, the phylogenetic tree exhibited only sequences that were the very high degree of similarity with SA76 from BlastP analysis. The emphasis was mainly focused on secreted aspartic proteinase from filamentous fungi, and the other typical sequences were also included in constructing the Phylogenetic tree. Fourteen sequences encoding secreted aspartic proteinase were selected and aligned using the ClustalX 1.81 program. A neighbour-joining tree was constructed using the Treeview. The phylogenetic analysis suggested that SA76 protein of *T. harzianum* clustered within *G. zeae, N. crassa, C. globosum* and *M. grisea*. The tree showed that *T. harzianum* shared the highest identity with *G. zeae*, followed by *N. crassa, C. globosum* and *M. grisea*. It shared the lowest identity with *C. immitis* as compared with the others.

Discussion

The BD SMART RACE technique is a newly developed PCR-based method for obtaining full-length cDNA. A complete 5' and 3' sequence of the target transcript can be isolated more consistently (Chenchik *et al*. 1995; Chenchik *et al*. 1996). Gene-Specific Primers plays an important role in RACE technique. All the primers should be 23–28nt long and primers longer than 30nt were generally no advantage. A necessary condition for GSPs was GC content of 50%–70% and Tm at least 65°C, whenever the Tm can be above 70°C. Longer primers with annealing temperatures above 70°C gave more robust amplification

in RACE. Tm over 70°C allows using "touchdown PCR". However, the use of self-complementary primer sequences should be avoided, which easily resulted in fold-back and formed intramolecular hydrogen bonds. Similarly, the use of primer complementarity in the Universal Primer Mix, particularly in their 3' ends should also be avoided. In 3' RACE, no specific band, only a smear was found in the products of the first PCR, while two bands were founded in the second PCR products. One was target band and the other was non-specific band verified by sequencing. It was inferred from this result that nest PCR was a useful method of RACE.

SA76 protein of *T. harzianum* is close to that of *G. zeae, N. crassa, C. globosum* and *M. grisea*. Most secreted aspartic proteinase from Monilales and Deuteromycotina fungi, also including those from *T. harzianum* and *Magnaporthe*, were in the same clade group. In contrast, *Gibberella* (Sphaeriales), *Neurospora* (Sphaeriales) and *Chaetomium* (Chaetomiales) belonged to Ascomycota. The classification of fungi was uncertainties, according to the typical characteristics some fungi belonged to Deuteromycotina whereas some fungi were classified to other subphylum. Thus phylogenetic analysis of SA76 gene with the closest secreted aspartic proteinases allowed us to identify its footing in evolving and to ensure the relationship with other fungi, consequently ascertain status of *T. harzianum.* The foundation was well-established to obtain and validate functional genes of *T. harzianum*.

The deduced amino acids shared 53% identity with secreted aspartic proteinase, a biocontrol protease gene of *G. zeae*, but no similar genes about *T. harzianum* were reported by the BLAST search. In our experiment, the secreted aspartic proteinase gene from *T. harzianum* was cloned for the first time. The cloning and analysis of secreted aspartic proteinase gene possibly provide theoretical support for researching its structures and expressions, and elucidate its functional mechanisms and the relationship with biocontrol of *T. harzianum* on the level of molecular.

References

Chenchik, A., Moqadam, F., Siebert, P. 1995. Marathon cDNA amplification: A new method for cloning full-length cDNAs. *Clontechniques,* **1:** 5–8.

Chenchik, A., Moqadam, F., Siebert, P. 1996. A new method for full-length cDNA cloning by PCR. In: Krieg, P. A (Ed), *A Laboratory Guide to RNA*: *Isolation, Analysis, and Synthesis*. Wiley-Liss, Inc., p273–321.

Geremia, R.A., Goldman, G.H., Jacobs, D.A., *et al*. 1993. Molecular characterization of the proteinase-encoding gene prb1, related to mycoparasitism by *Trichoderma harzianum*. *Mol. Microbiol.*, **8**: 603–613.

Harman, G.E. and Björkman, T. 1998. Potential and existing uses of Trichoderma and Gliocladium for plant disease control and plant growth enhancement. *Enzymes, Biological Control and Commercial Application*, **2**: 229–265.

Liu Pigang, Yang Qian. 2005. Identification of genes with a biocontrol function in *Trichoderma harzianum* mycelium using the expressed sequence tag approach. *Research in Microbiology*, **15**: 416–423.

Louie, K., Patricia, A.C. 1999. Characterization of a cDNA encoding a subtilisin-like serine protease (NC-p65) of *Neospora caninum*. *Molecular and Biochemical Parasitology*, **103**: 211–223.

Manczinger, L., Polner, G. 1985. Cluster analysis of carbon source utilization patterns of Trichoderma isolates. *Syst. Appl. Microbiol.*, **9**: 214–217.

Nathalie, P., Stephanie, G., Christine, R., *et al*. 2001. AspS encoding an unusual aspartyl protease from Sclerotinia sclerotiorum is expressed during phytopathogenesis. *FEMS Microbiology Letters*, **194**: 27–32.

Schaefer, B.C. 1995. Revolutions in rapid amplification of cDNA ends: new strategies for polymerase chain reaction cloning of full-length cDNA ends. *Analytical Biochemistry*, **227**: 255–273.

Zheng Ye, James, R.C. 2000. cDNA cloning by amplification of circularized first strand cDNAs reveals non-IRE-regulated iron-responsive mRNAs. *Biochemical and Biophysical Research Communications*, **275**: 223–227.